Visualizing Social Science Research in an Institutional Repository

Ted Polley
Social Sciences Librarian
University Library
Indiana University-Purdue University Indianapolis

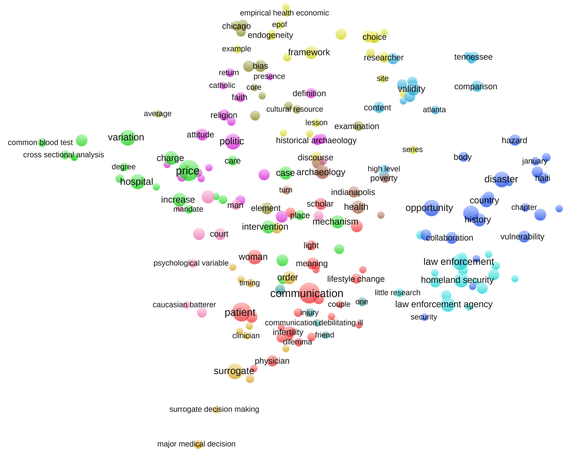

Figure 1. Term map of social science research archived in IUPUI ScholarWorks. Colors represent different areas of research.

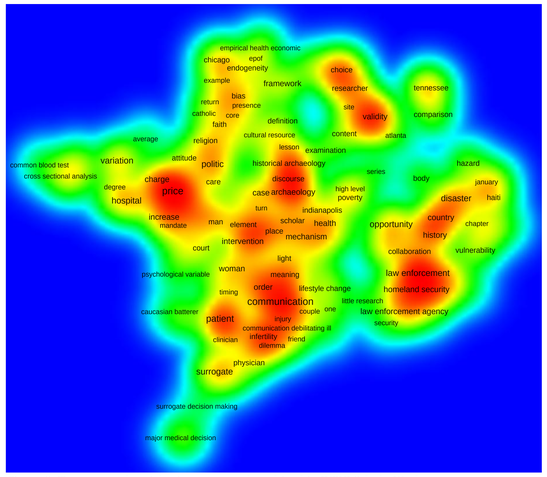

Figure 2. Term map of social science research archived in IUPUI ScholarWorks. Colors represent density of terms.

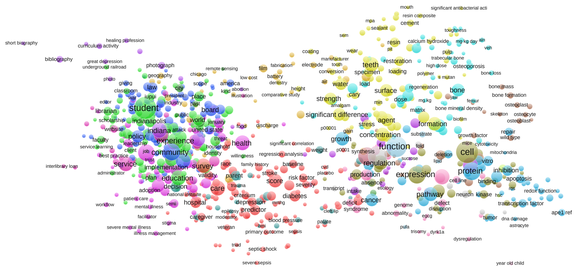

Figure 3. Term map of all research archived in IUPUI ScholarWorks. Colors represent areas of research.

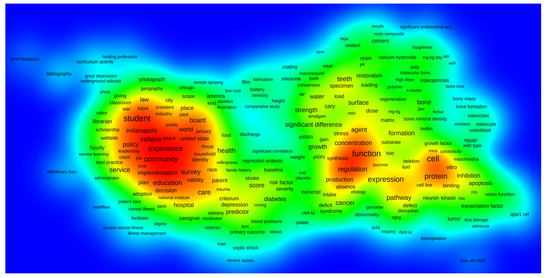

Figure 4. Term map of all research archived in IUPUI ScholarWorks. Colors represent density of terms.

Table 1. Term frequency, relevance, and excluded terms for social science term map.

20 Most Frequently Occuring Terms

Term	Occurrences	Relevance
price	20	0.67
communication	20	0.45
patient	17	0.5
law enforcement	13	0.7
opportunity	13	0.37
surrogate	12	1.05
variation	12	0.84
politic	12	0.8
disaster	12	0.56
recommendation	12	0.37
hospital	11	0.83
law enforcement agency	10	0.93
homeland security	10	0.66
woman	10	0.51
archaeology	9	0.83
charge	9	0.62
validity	8	1.21
order	8	0.72
framework	8	0.65
history	8	0.56

20 Most Revelevant Terms

Term	Occurrences	Relevance
stage estimator	3	4.84
older adult	3	2.85
empirical health economic	3	2.67
major medical decision	4	2.44
chicago	5	2.26
preservation policy	5	2.15
caucasian batterer	3	2.06
catholic	3	2.03
measurement	3	1.92
everyday life	3	1.92
price effect	5	1.89
return	4	1.87
cultural district	3	1.8
thing	5	1.73
factorial	3	1.6
initial validation	5	1.57
internal consistency reliability	6	1.53
atlanta	4	1.51
common blood test	4	1.49
memphis	4	1.48

Excluded Terms

Term	Occurrences	Relevance
method	20	0.56
conclusion	16	0.48
instrument	11	1.46
increase	9	0.72
review	8	1.11
scope	3	0.74

Table 2. Term frequency, relevance, and excluded terms for full ScholarWorks term map.

20 Most Frequently Occurring Terms

Term	Occurrences	Relevance
student	375	1.04
cell	375	0.75
function	333	0.44
mechanism	280	0.39
disease	248	0.29
protein	245	0.78
expression	237	0.63
community	224	0.77
practice	214	0.87
experience	209	0.79
indiana	178	0.93
gene	174	0.68
health	170	0.49
education	168	0.77
service	167	0.96
care	166	0.57
mouse	164	0.78
teeth	157	1.14
resource	156	0.74
survey	155	0.68

20 Most Relevant Terms

Term	Occurrences	Relevance
short biography	10	8.63
healing profession	19	5.22
cleft lip	20	4.78
rosenstein	13	4.35
conscientious objection	23	4.2
palate	22	4.1
bibliography	32	3.4
mouth	11	3.04
general practitioner	15	2.78
dental pulp	16	2.75
farm security administration	12	2.75
calcium hydroxide	21	2.6
severe sepsis	13	2.5
sealant	26	2.46
septic shock	22	2.41
open access material	14	2.34
curriculum activity	26	2.26
significant antibacterial activity	10	2.26
pediatric dentist	13	2.22
winter	12	2.15

Excluded Terms

Term	Occurrences	Relevance
article	163	0.92
author	69	0.95
book	55	1.21
bulletin	189	10.01
essay	21	1.26
figure	15	0.78
health bulletin	24	10.61
indiana state board		
investigator	38	0.66
issue	195	0.7
journal	38	1.04
newsletter	43	2.54
poster	35	0.87
publication	34	0.78
research project	30	0.73
vol	259	8.99

The Data

IUPUI ScholarWorks serves as IUPUI's institutional repository, reflecting the research and creative output of the university. The repository runs on DSpace and hosts 25 different research communities. In February 2015, community-level metadata, containing standard Dublin Core elements for each item, were exported to CSV files. For this project, the dataset was limited to collections likely to include social science research:

- Faculty Articles (a multidisciplinary collection of faculty research)
- School of Liberal Arts
- School of Social Work
- School for Public & Environmental Affairs (SPEA)

Data Preparation and Visualization

The first step in cleaning the data required limiting the Faculty Articles and School of Liberal Arts collections to only research that can broadly be categorized as social science research. There are multiple methods for limiting to just social science research, but the most feasible for this project relied on authors' departmental or school affiliation. A list of faculty in the anthropology, communication studies, economics, political science, and sociology departments was combined with a list of faculty in the School of Social Work and SPEA. The removal of non-social science items was done in Excel. Each CSV file was then loaded into R and the title, abstract, and item ID were extracted. These variables were combined into a new dataset and de-duplicated using the item IDs (an item's membership in a community is not mutually exclusive).

The resulting dataset consists of 154 unique items. The titles and abstracts for each item were written to a text file and visualized using VOSviewer. This tool is commonly used for bibliometric analysis, but is well suited for large-scale textual analysis.[1] Term co-occurrence maps were generated for the social science research dataset and the entire ScholarWorks collection for comparison.

Results

The social science dataset is a relatively small amount of data for large-scale textual analysis, especially when compared with the 5066 titles and abstracts in the full ScholarWorks dataset. Regardless, there are still insights to gain from these term maps. Unsurprisingly, there is a prominence of health-related research in both maps, as IUPUI is a major health sciences campus.

In the social science term map a few key faculty/researchers who are heavy contributors to ScholarWorks force the map to coalesce around health communication research, health economics, archaeology, and natural disaster research, evident in the clusters (Figure 1) and the term densities (Figure 2). Unexpectedly, the area of highest relative density in the full ScholarWorks dataset term map is in the social sciences and humanities (Figure 4). This is likely due to a volume of research across various disciplines that focuses on student learning and community engagement. It is also likely due to a large collection of the Geography Educators' Network of Indiana Newsletter that is stored in the repository.

References

[1] Van Eck, N. J., & Waltman, L. (2011). Text mining and visualization using VOSviewer. arXiv:1109.2058 [cs]. Retrieved from http://arxiv.org/abs/1109.2058